Theoretical, Empirical, and Computational Perspectives on Hungarian Discourse Configurationality

by Hannah Gendler Szabó

A thesis for the Bachelor of Science in Computing and Linguistics

Pauli Murray College
Yale University

Advised by Tom McCoy and Robert Frank

25 April 2025

Acknowledgments

I am deeply grateful to have been guided through this project by both Tom McCoy and Bob Frank, who in weekly meetings attended to various aspects of this project, from the most minute to the more general, each in their own characteristically kind way. I am also indebted to other professors in the Linguistics department, namely Larry Horn, Jason Shaw, Raffaella Zanuttini, Jim Wood, and Veneeta Dayal (who taught me in my first-ever linguistics class and piqued my interest in the major). Within the department, thanks also to Miranda Zhu, whose research inspired my own and who offered up insights into her own methodology, and to Samuel Ostrove and William Min, the other two inaugural Computing and Linguistics majors, for company and support.

For the financial support to undertake an empirical survey, I am thankful to the Pauli Murray College Mellon Research Grant, and to my Head of College Tina Lu. The Yale Center for Research Computing provided access to high-performance systems that allowed me to undertake the computational portion of this project. And I am very thankful to the fifty anonymous survey-takers for sharing their intuitions about language with me.

Lastly, I would like to thank the instructors who taught me Hungarian over the years. My babysitter in infanthood, Katalin Laczkó, taught me my first words, which were in Hungarian. Nineteen years later, Szilvia Papp, Valeria Varga, and Carol Rounds re-taught me the language with care and creativity. I am especially grateful to Valeria for inspiring a deep love of the language and its syntactic quirks. Lastly, I dedicate this thesis to my grandparents, Zsuzsanna Olti and László Szabó, who spoke to me affectionately in the language throughout my childhood. It was for their sake that I learned Hungarian, and I am grateful to have had the opportunity to explore the fascinating language from so many angles.

Theoretical, Empirical, and Computational Perspectives on Hungarian Discourse Configurationality

Abstract Hungarian employs complex, contextually-sensitive syntactic rules that place certain constituents in the pre-verbal positions of Focus and Topic. This thesis examines this property of Hungarian, known as discourse configurationality, by comparing native speakers' grammaticality judgments of possible word orderings within a certain context to the probabilities language models assign in the same situations. After reviewing the discourse-semantic and syntactic aspects of Hungarian word order, I introduce a basic taxonomy of Hungarian focusing and topicalization rules based on foundational perspectives from generative linguistics, and construct a novel set of Hungarian scenarios testing all possible intersections of these rules (5×4). I assess the grammatical judgments of native speakers (n=50) through a survey administered on the online platform Prolific. These results serve as both a quantitative evaluation of the descriptive capacity of the syntactic rules posited by linguists, as well as the baseline against which the performance of SambaLingo-Hungarian-Chat (Llama-2-7b, optimized for Hungarian) is evaluated. By examining the capability of language models to replicate native-like word order in Hungarian, I present a case for using word-order predictions of discourse configurational languages as a diagnostic for assessing the ability of language models to recognize and respond to subtle pragmatic details, and more broadly for the importance of utilizing insights from theoretical linguistics to engineer better diagnostics for language model evaluation.

Introduction

As an agglutinative language, Hungarian encodes grammatical function through case-marking affixes. These affixes allow for more variance in word order, unlike languages like English which lack such inflectional markers and rely on fixed syntactic positions to maintain interpretability. Although Hungarian word order is sometimes referred to colloquially as 'free,' word order is used to convey important discursive information through the movement of constituents into two pre-verbal positionings known as Topic and Focus. Because Hungarian syntactic organization reflects discourse-semantic roles, it becomes an interesting case study to explore the relationship between syntax and pragmatics.

Linguists studying word order in Hungarian have developed various theories of how constituents are selected for Topicalization and Focusing—more precisely, how particular constituents acquire the features [TOPIC] and [FOCUS] that motivate their movement to pre-verbal positions. In this vein, my research aims to examine what patterns of association exist between discourse roles and constituent order in Hungarian, and to empirically quantify the consistency of these patterns through a survey of native Hungarian speakers. In addition, my thesis explores the ability of language models to replicate the sort of discourse-sensitive word ordering used by human speakers. By comparing syntactic theory, human grammaticality judgments, and the probabilistic outputs of language models, I aim to measure how formal accounts, subjective judgments, and machine-generated predictions describe the same linguistic structures.

My paper has five parts: (§1) Linguistic Perspectives; (§2) Dataset and Experiment Description; (§3) Survey Results and Discussion; (§4) Language Model Comparative Results and Discussion, and (§5) Conclusion and Further Directions. In §1, I discuss the syntax,

semantics, and pragmatics of Hungarian pre-verbal positions, and sketch out previous attempts to articulate selection rules for Focusing and Topicalization in Hungarian. §2 introduces a dataset of 100 Hungarian context-response sentence pairs distributed across an introduced taxonomy of Focusing and Topicalization, and lays out my methodology for testing native Hungarian speakers and optimized language models on this data. In §3-4 I present the results of both my empirical survey and my language model evaluations, which provide important insights separately and in comparison to one another. Lastly, §5 considers the assumptions that underlie my experimentation, the implications and limitations of my findings, and proposes further avenues for research.

§1: Linguistic Perspectives

1.1 Syntax of Focusing and Topicalization

The foundational theory of the syntax of Focus within generative grammar comes from Jackendoff (1972), who introduces "a syntactic feature F which can be associated with any node in the surface structure" that causes the "associated semantic material [to be] the Focus of the sentence" (Jackendoff 1972: 240). Topicalization is likewise assumed within the minimalist framework to be a feature-based movement of particular constituents to a position that c-commands the verb phrase (É. Kiss 2002, 12).

Horváth adapts Jackendoff's general framework to "Hungarian-type languages," first in her 1981 dissertation, and in expanded form in her 1985 book. Horváth notes that to explain the syntactic movement associated with Focus, V must also carry a [FOCUS] feature which triggers the movement of the [FOCUS]-marked constituent to the preverbal position (Horváth 1985: 132).[1] Focus movement is thus analogous to movement motivated by Case assignment.

Generalizing the theory Horváth puts forth about Hungarian syntax, É. Kiss (1995) formalizes the category of discourse configurational languages, as those wherein constituents move into particular syntactic positions in order to represent their discursive roles. In other words, movement in discourse configurational languages tends to be the result of agreement between [FOCUS] and [FOCUS]-like features rather than Case features. É. Kiss introduces two particular positions that appear cross linguistically: the Topic position ("A") and the Focus position ("B"). Hungarian belongs to "type AB" according to this taxonomy, meaning it utilizes

[1] Focusing also typically causes the focused constituent to receive prosodic stress, while topicalized constituents cannot typically receive stress. The prosodic aspects of focusing and topicalization have been examined extensively, for example by Zubizarreta (1998) crosslinguisticaly, and by Sendrői (2017) for the specific case of Hungarian. For the purpose of this analysis, which focuses on *written* Hungarian sentences, the role of prosodic stress will be largely set aside.

both possible pre-verbal discursive positions.[2] In Hungarian, the Topic position is the specifier of the Tense Phrase (TP), c-commanding the Verb Phrase (VP), while the Focus position is the specificer of the Verb Phrase. Thus, Focused constituents always appear in the immediate pre-verbal position while Topicalized constituents always appear before Focused constituents.

In line with Horváth's observations, É. Kiss (2002) proposes that Hungarian Focusing is a feature-motivated process analogous to the Extended Projection Principle (EPP), wherein the [FOCUS] feature on the verb must be 'checked' by being c-commanded by a constituent also carrying [FOCUS] (É. Kiss 2002: 89). In Hungarian specifically, the Topic position is the specifier of the Tense Phrase (TP), c-commanding the Verb Phrase (VP), while the Focus position is the specificer of the Verb Phrase. Effectively, this theory of focusing encodes the requirement that Spec-v*P must be occupied by a constituent carrying the [FOCUS] feature. Similarly for the case of topicalization, an EPP-like feature is posited that triggers A-bar movement to the left periphery of the TP (analogous to the movement of the grammatical subject in English out of v*P).

Consider sentence (1), which has a topicalized subject and a focused object. Underlining marks the Focused constituent. [Brackets] enclose the Topicalized constituent. A syntax tree portraying the underlying structure and movements is included below.

(1) [János] az imá-t mond-ja
John the prayer-ACC say.PRS.3SG-DEF
'[John] says the prayer' / 'It is the prayer that [John] says'

[2] Other examples of discourse configurational languages include Basque, Catalan, Russian, Korean, Yoruba, and Quechua. Japanese is the paradigmatic example of a "type A" discourse configurational language, and Aghem is "type B."

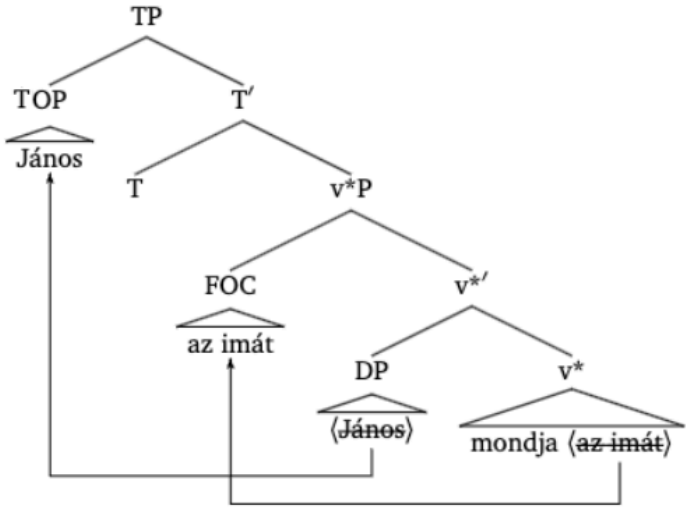

Figure A: syntactic structure of (1)

In the above example, [$_{DP}$ *az imát*] moves from v* (where it acquires its case marking) into the Focus position in order to check the [FOCUS] feature with which it entered the syntax against the analogous feature carried by v*'. Likewise, [$_{DP}$ *János*] moves from DP to the Topic position to check a [TOPIC] feature. This theory of the syntax of Focus and Topicalization offers an explanation for why Focused constituents always appear in the immediate pre-verbal position while Topicalized constituents always appear before Focused constituents.

This positioning of the Focused constituent can be verified by a quantifier scope diagnostic. In Hungarian, constituents that move to the left-peripheral focus position ([Spec,FocP]) take scope over elements to their right, including in-situ quantifiers. Consider the contrast between the following two sentences:

(2) a. János egy imá-t mond mindenki-nek.
János a prayer-ACC say.PRS.3SG.IND everybody-DAT

'János says a prayer to everyone' (There is one prayer that everyone heard)

b. János mindenki-nek mond egy imá-t
János everybody-DAT say.PRS.3SG.IND a prayer-ACC

'János says a prayer to everyone' (He may have said a different prayer to each)

In (2a), the focused direct object *egy imát* takes scope over the quantifier *mindenkinek*, yielding the interpretation that there was one particular prayer that was recited to all. In contrast, (2b), where *egy imát* appears postverbally, the scope relation is reversed: each person may have heard a different prayer. This contrast shows that Focus movement to [Spec,FocP] imposes a scope-rigid reading, confirming the structural position of the Focus position within the syntax.

1.2 Semantics and Pragmatics of Focusing and Topicalization

Grice's Maxim of Relevance, part of his Cooperative Principle, emphasizes that speakers should ensure their contributions are pertinent to the ongoing conversation (Grice 1975). This principle suggests that any deviation from the current topic requires clear signaling, as unmarked shifts can lead to confusion or misinterpretation. The intuitive purpose of topicalization, as means of introducing a new discourse-entity to be predicted upon, aligns with this principle (Rounds 2009).

In her paper introducing the terminology of Question Under Discussion (QUD), Roberts (1998) draws from Grice (as well as Stalnaker and early thinkers in artificial intelligence), framing discourse as "organized around a series of *conversational goals* and the plans or *strategies* which conversational participants develop to achieve them." Roberts suggested that an utterance is relevant *iff* it addresses what she calls the QUD. Following Roberts (1998), Beaver et al (2017) define the syntactic Focus position as the part of an utterance as "that part of the utterance that answers the QUD." In line with these pragmatic understandings of focus, Jackendoff (1972) sees Focus as the component whereby the assertion differs from the presupposition.

1.3 Distribution of Hungarian Focus and Topic Features

Having established these properties of focusing and topicalization, I now turn to the question of what principles govern the distribution of the [FOCUS] and [TOPIC] features in Hungarian. Here, my aim is not to select one theory over others, but to outline a set of observations made by linguists about the sorts of selectional tendencies that govern which constituents are assigned these features.

1.3.1 Distribution of the Focus Feature

The most obvious observation to make is that Wh-words always enter the syntax with a [FOCUS] feature. Horváth (1985) proposes an "essential parallelism" between Wh-movement and Focusing in Hungarian, and suggests that there is no need for an independent explanation of Wh-movement beyond a special example of Focusing (118).[3] Likewise, É. Kiss (1995) suggests that the constituents that directly answer Wh- questions also carry [FOCUS], a view Horváth (2005) takes up as well. In the case of Wh- questions and their answers, keeping these constituents *in situ* or moving them elsewhere other than the preverbal position results in ungrammaticality, as in the sentences below:

(3) a. {Ki-t} lát-ott {*ki-t} Mari {*ki-t}?
*{who-ACC} see-PST.3SG.IND {*who-ACC} Mari {*who-ACC}*

'Whom did Mary see?'

b. {*Miki-t} [Mari] {Miki-t} lát-ta {*Miki-t}.
*{Nick-ACC} Mari {Nick-ACC} see-PST.3SG.DEF {*Nick-ACC}*

[Mary] saw Nick.

[3] É. Kiss (2002) also follows this formulation of Wh-movement as an example of Focus movement. For a counterperspective, see Seth Cable's 2008 manuscript "Fronting (in Hungarian) is Not Focus-Fronting."

We can also characterize the distribution of [FOCUS] by considering two additional discourse contexts in which constituents tend to bear the feature. First, newly introduced referents or propositional elements—those that have not previously been accessible in the discourse—frequently appear in the preverbal focus position. This aligns with the well-documented correlation between Focus and new information (see Lambrecht 1994, É. Kiss 1995). Second, constituents involved in the clarification, contrast, or specification of a relationship between two entities also tend to be focalized. This is particularly evident in cases where the Focus-marked element serves to update or refine a presupposition. The following examples demonstrate these two situations, where *b*. is a response to *a.*:

(4) a. Képzeld, előléptet-tek valaki-t!
imagine.IMP promote-PST.3PL.IND someone-ACC

'Guess what, someone got promoted!'

b. <u>Péter</u> kap-ta az előléptetés-t.
Peter receive-PAST.3SG.DEF the promotion-PST.INDF.3PL

'<u>Peter</u> got the promotion.'

(5) a. A gyerek és a kutya játsz-ott-ak.
the child and the dog play-PST-3PL

'The child and the dog played.'

b. [A gyerek] a <u>kutya-val</u> játsz-ott-∅?
the child the dog-INS play-PST-3SG

'Did [the child] play <u>with</u> the dog?'

To formalize these particular trends in Focusing, I introduce two features: [NEW], marking discourse-new elements, and [REL], marking an updated relationship between two discourse-old elements. These distribution characteristics align with the pragmatic claims summarized in §1.2.

1.3.2 Distribution of the Topic Feature

It is well-attested that temporal phrases seem to enter the syntax with the [TOPIC] feature, which explains their frequent appearance in sentence-initial position (see Rounds, 2009). As noted in §1.1, discourse configurational languages have a topic-predicate structure, rather than a subject-predicate structure, so the Topicalized element is not necessarily the grammatical subject. É. Kiss notes that "the link between subjecthood and topichood is only indirect" in Hungarian, and thus that [NOM] (the feature that represents nominative case) does not necessarily co-occur with [TOPIC]. Specifically, É. Kiss argues that [NOM] is a weaker predictor of Topicalization in Hungarian than [HUMAN] (2006: 9). This would imply the for the following sentences, the "i" options occur with more frequency than the "ii" options:

(6) i. János-t meg-harap-ja a kutja
John-ACC COMPL-bite-PRS.3SG the dog

'János was bitten by the dog'[4]

ii. A kutja meg-harap-ja János-t
the dog COMPL-bite-PRS.3SG John-ACC

'The dog bites János'

(7) i. János meg-harap-ja a kutjá-t
John COMPL-bite-PRS.3SG the dog-ACC

'János bites the dog'

ii. A kutjá-t meg-harap-ja János
the dog-ACC COMPL-bite-PRS.3SG János

'The dog was bitten by János'

These observations leave us with four relevant features to consider in relation to Focusing ([WH-Q], [WH-ANS], [NEW], [REL]), and three in relation to Topicalization: ([TEMP], [NOM], and [HUM]).

[4] Note that Hungarian has no passive voice constructed with the use of auxiliary verbs and past participles, and instead makes use of the topic-focus construction as making use of the verb *megvan* ("to be done").

§2: Dataset and Experiment Description

2.1 Dataset

I now introduce a dataset of 100 assertion-response pairs constructed specifically to involve both Focusing and Topicalization. These pairs will then be used in order to test word order preferences (in the case of native speaker evaluations) and probability assignments (in the case of language model evaluations), to allow for comparison between human intuition and computational predictions.

Each pair contains a single assertion — either a question or a statement. This assertion can be understood as the preceding statement made by the interlocutor. The response is made up of three constituents: a verb (*V*), and two non-verbal constituents (*a* and *b*) which may or may not be moved to Topic and Focus positions (referred to as *V*, *a*, and *b* hereafter). This yields six possible constituent orderings for the response, depending on whether or not *a* and *b* are Topicalized and Focused:

1) *a* | *b* | *V*
2) *b* | *V* | *a*
3) *a* | *V* | *b*
4) *b* | *a* | *V*
5) *V* | *b* | *a*
6) *V* | *a* | *b*

To systematically test the role of different features in these movement processes, I ensured that all the features introduced in §1 are evenly represented across the dataset. In each response, *a* and *b* are assigned different Focus-related and Topic-related features in order to examine to what extent sentences with constituents carrying similar features are treated similarly. Because the features [NOM] and [HUMAN] are not mutually exclusive, I combine these features to examine their interaction: [NOM + HUMAN]. I also consider a (possibly Topicalized) constituent that

carries none of the features under consideration, marked with [—]. With four Focus-related categories and five Topic-related categories, this yields a total of twenty possible sentence categories:

F-features	wh_q	wh_ans	new	rel
T-features				
nom				
hum				
nom+hum				
temp				
—				

Figure B: chart depicting possible sentence categories based on distribution

Each category thus contains five assertion-response pairs to minimize the risk of overfitting to specific lexical items and ensure that patterns observed are due to syntactic and discourse features rather than the particular words used in the sentences.

2.2 Experiment A: Human grammaticality judgments

To assess native speaker grammaticality judgments, I built a survey using the web-based survey tool Qualtrics XM, and administered it to native Hungarian speakers through the online platform Prolific. The survey was completed by 50 native Hungarian speakers, whose eligibility was determined via Prolific's language screening feature. I required participants to list Hungarian as both their native and their current primary language. Before full data collection, I conducted a test run with 10 participants to evaluate the clarity of the survey design and question format. These test responses were excluded from the final dataset, as minor revisions were made

afterward, including correcting one definite noun to an indefinite noun and ensuring that all response fields were mandatory to prevent missing data.

Each participant's responses were manually screened to ensure data quality, including checks for completion time and engagement levels. Participants were compensated according to an assumed $12 per hour rate for a 20-minute survey. The actual average completion time was 15.5 minutes, meaning participants earned approximately $14.48 per hour. Test survey participants were also compensated accordingly.

Each participant encountered 20 sentences, with each sentence representing one of the possible feature combinations to ensure that each participant saw each combination once. The sentences were presented in a randomized order to prevent ordering effects. For each sentence, participants were provided an initial assertion, and provided two types of grammaticality judgments about possible responses:

- Absolute Judgment: Participants rated the acceptability of each of the six possible word orders on a 0–4 scale, answering the following question:

  Mennyire találja elfogadhatónak a következő szórendet ebben a kontextusban?

  "How acceptable do you find the following word order in this context?"

- Relative Judgment: Participants then selected the most appropriate word order from the six options, answering the following question:

  Melyik szórend tűnik a legmegfelelőbbnek itt?

  "Which word order seems the most appropriate here?"

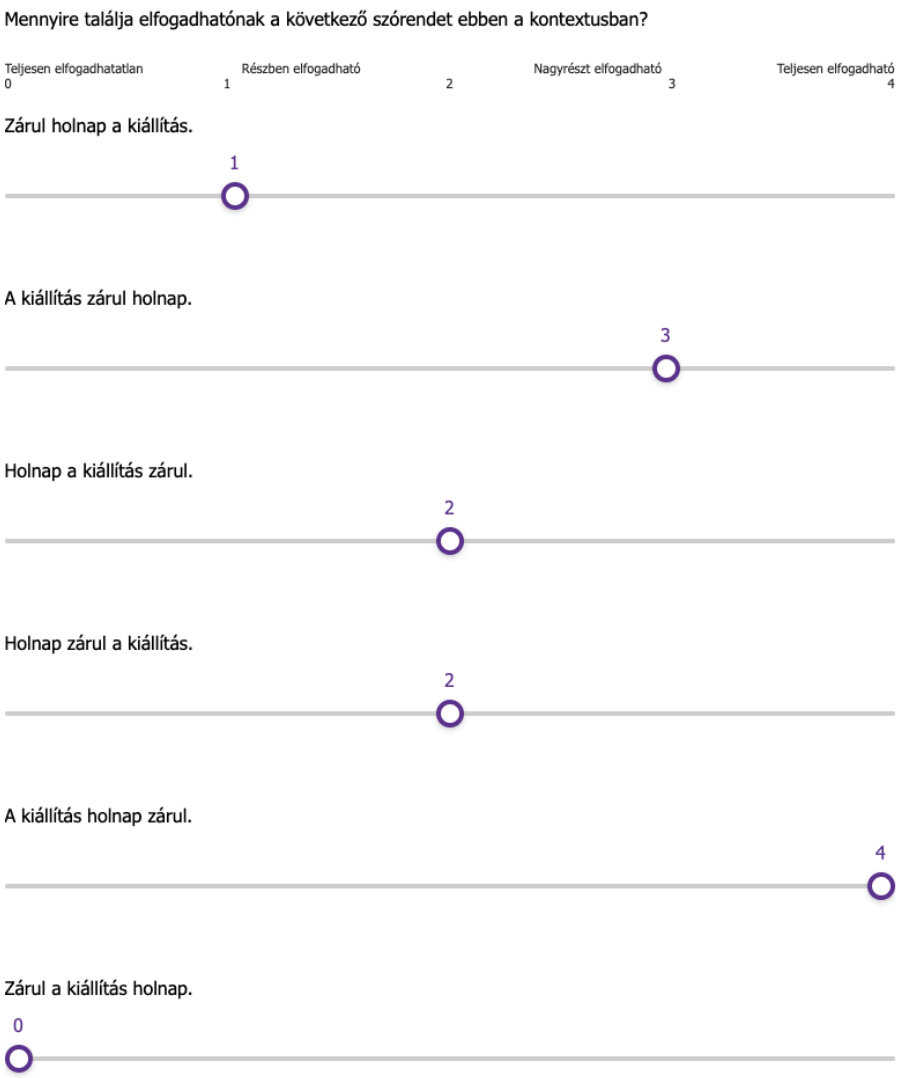

Figure C: example of response for an absolute judgement

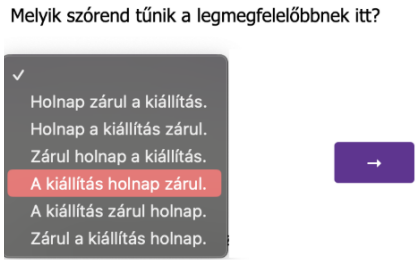

Figure D: example of response for a relative judgement

Including both absolute and relative judgments in the survey serves several purposes. First, it allows for a cross-checking mechanism to ensure participants are rating consistently and appropriately. Second, it provides a tie-breaker in cases where multiple word orders receive the highest absolute rating.

2.3 Experiment B: Language model word order probabilities

To compare native speaker judgments with computational predictions, I used the same dataset to evaluate SambaLingo-Hungarian-Chat, a human-aligned Hungarian-language chat model developed in 2024 by a team of multilingual-LM scholars at SambaNova. This model was trained on top of SambaLingo-Hungarian-Base, which is an adaptation from Llma-2-7b trained 59 billion tokens from the Hungarian split of the Cultura-X dataset. I ran the model on Grace, a shared-use resource within the Yale Research Computing Cluster. The model was accessed via Hugging Face's Transformers library, using standard inference settings.

Each model trial followed a structured procedure:

- A prompt was provided, identical to the “assertion” of the human survey questions
- Each of the six possible word orders was supplied as a potential "answer" to the question.
- The log probability of each answer given the question was computed by summing the token-level log probabilities assigned by the model.
- Since all possible answers contained identical tokens in different arrangements, log probabilities could be normalized by first converting them to probabilities and then ensuring they summed to 1, yielding a probability distribution over the six possible word orderings.

This approach ensures that probability assignments reflect only differences in syntactic preferences rather than lexical variation, allowing for direct comparison with native speaker grammaticality judgments.

§3: Survey Results and Discussion

In the following section, I present and discuss the results of the survey. I summarize the trends that emerge from the aggregation of the relative ratings (that is, of the selected 'best' word orderings) across the taxonomy introduced in §2.1. Next, I examine the absolute ratings to provide further nuance on the findings of the previous section.

3.1 Human survey, relative ratings

At the highest level, the human survey results (see following page) reveal a difference in the degree of optionality between Focus-marking and Topicalization. Across all questions, participants selected either the *abV* or *bVa* word order—both of which involve Focus on the *b* constituent—in 74.9% of cases (95% CI: ±0.08). However, when considering more specific alignment with both Focus and Topicalization cues, only 26.6% of participants selected the *abV* order, which exactly matches the predicted structure. This suggests that while Focus-marking is relatively stable, the additional pressure to Topicalize is more optional or less consistently followed across discourse contexts.

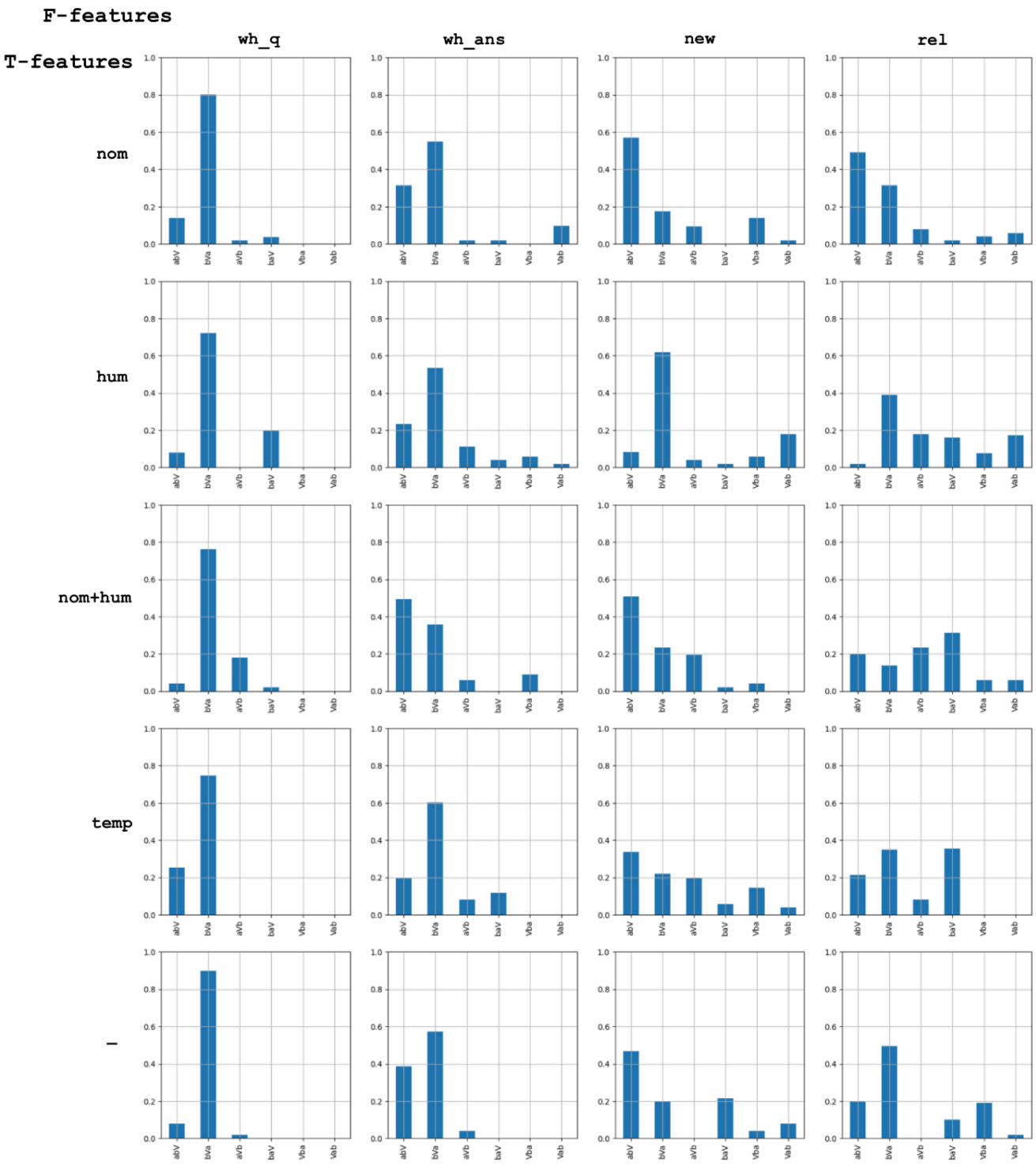

Figure E: aggregated selections for top word ordering by scenario category

The results reveal a strong preference for Focusing the Wh-question (see the second bar in each graph in the leftmost column). For sentences that *answer* a Wh-question (see the second column

from the left), the preference remains very strong to Focus constituent *b*, and there is more variance as to whether or not *a* is Topicalized as compared to sentences that contain a Wh-question. For the two Focus-related properties that do not have to do with Wh-questions, `rel` and `new`, answers that Focus constituents carrying these features are still most commonly picked, but the difference is notably more moderate (note the greater spread in selections in the rightmost two columns).

There appears to be a stronger preference for Topicalizing attributes marked with `nom` than those marked with `hum`. Topic-initial structures were selected more often in `nom`-marked contexts (M= 0.32) than in `hum`-marked contexts (M = 0.17), although a paired-samples t-test found that this difference was not statistically significant ($t(4) = 0.81$, $p = .461$). Sentences in which constituent b carries both `nom` and `hum` appear more similar to those carrying in which b carries only nom than those in which it carries only `hum`: the mean value of the abV column for the `nom+hum` row is .38, closer to the corresponding mean of the `nom` row (.31) than that of the hum row (.10). This suggests that `nom` may be a more dominant or salient feature for determining topicalization than `hum`, and that grammatical features associated with definiteness or referentiality (like nominative marking) play a stronger role in licensing topicalized positions than semantic features like humanness. Interestingly, there is no clear evidence of a preference for Topicalizing temporal expressions in sentences of this length: the mean value of the *abV* column for the `temp` row is .25, which is in fact slightly lower than the un-featured "`-`" row (.28).

Taken together, the column-wise structure of the graphs (grouped by Focus-relevant features) reveals more consistent trends than the row-wise structure (grouped by Topic-relevant features). Quantitatively, the variance in the selection of the Focus-initial structure across

columns (0.0164) was lower than the variance in the selection of the Topic-initial structure across rows (0.0267), though this difference did not reach statistical significance ($t(8) = 0.91$, $p = 0.39$). Nevertheless, the numerical trend supports the interpretation that Focus-relevant features elicit more consistent ordering preferences than Topic-relevant features.

3.2 Human survey, absolute ratings

The absolute judgment results largely confirm the findings of the relative judgments, but they also clarify the difference between judgments made about Wh- questions and answers to Wh-questions (see Figure F on the following page). Although the relative judgment demonstrated that both Wh- questions and their corresponding answers are judged as needing to be in the Focus position, violating of this rule is much less acceptable for Wh- questions than for the corresponding answers. Interestingly, even though for the relative judgments, the word orderings that Focus the constituent with Focus-relevant features are always preferred, the margin for the absolute judgments of these options over the others is very small or sometimes nonexistent.

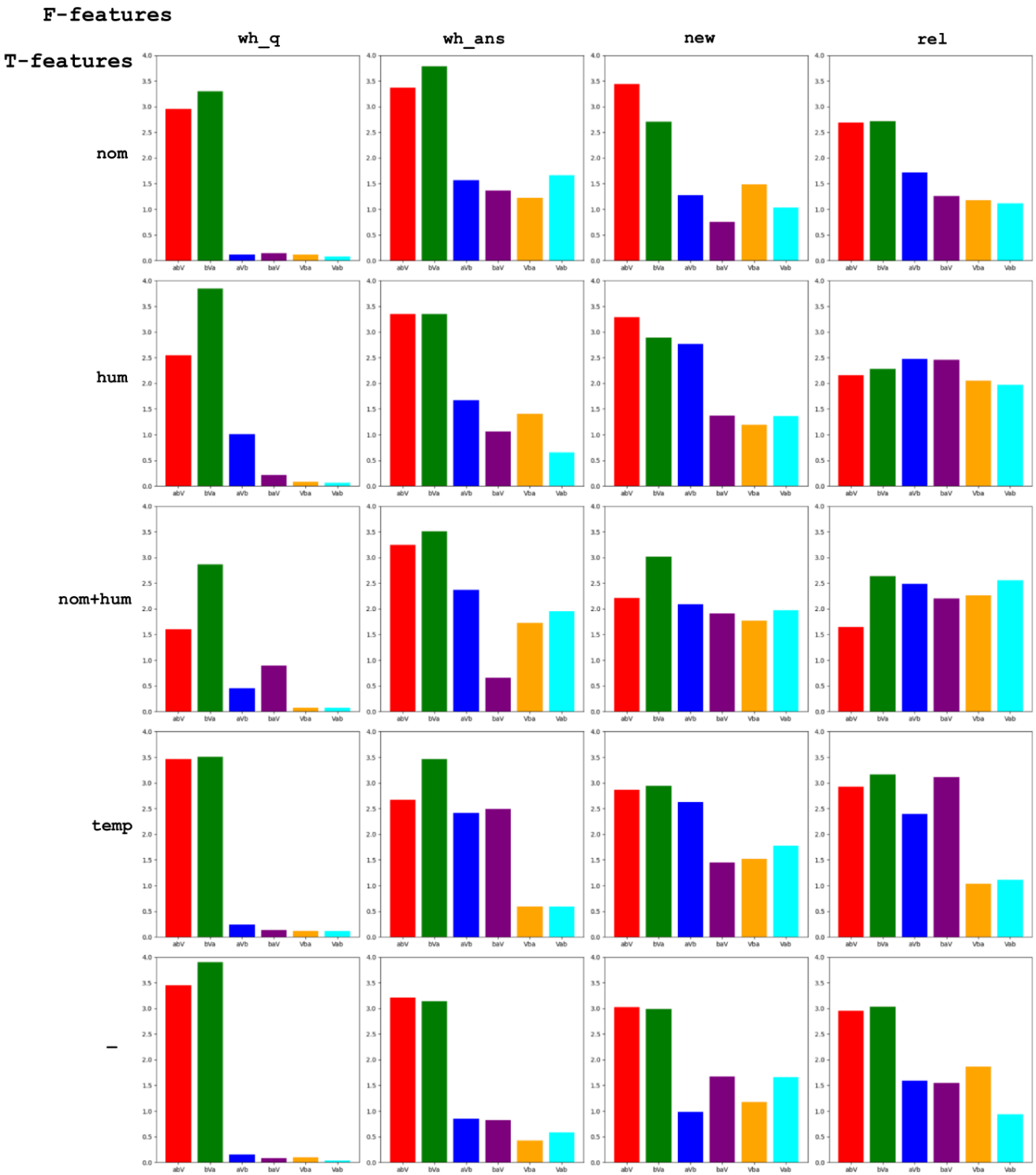

Figure F: average acceptability judgment given to each possible word ordering, by scenario category

§4: Language Model Comparative Results and Discussion

I now introduce the probability assignments made by the language model for each word order scenario and compare them to the human survey results presented in §3. I provide both a descriptive and quantitative assessment of the degree to which the language model's preferences align with human judgments, using KL-divergence as a measure of distributional divergence.

4.1 Qualitative observations

In several cases, especially where human participants exhibit a strong and consistent preference for a particular word order (see Figure F on the following page, column 1, row 1), the language model assigns its highest probability to the same word order, suggesting partial alignment between model predictions and human intuitions. These points of convergence are most prominent in contexts where discourse cues are particularly strong and unambiguous — most notably in the `wh-q` and `wh-ans` conditions (see columns 1 and 2).

In some cases, particularly where there is a strong preference for a specific word order among human participants (e.g., the first row, first column), the language model also assigns a higher probability to that same word order. However, there are instances where the model distributes probability more evenly across multiple word orders compared to the strong preference observed in human judgments (see row 4, column 2; and row 1, column 3). There are also notable outlier cases — such as row 1, column 4 and row 3, column 3 — where the reverse is true: the language model exhibits a strong preference for a specific ordering, while human responses are more distributed.

When comparing across feature conditions, clear patterns again emerge. In the case of `wh-q` and `wh-ans` sentences (columns 1 and 2), model outputs are generally more aligned

with human judgments, especially in rows where human preferences are strong. By contrast, in the `new` and `rel` conditions (columns 3 and 4), discrepancies between model and human judgments are more pronounced. This suggests that the model struggles more with interpreting pragmatic nuances that underlie novelty or information updating when not overtly introduced as the answer to a question.

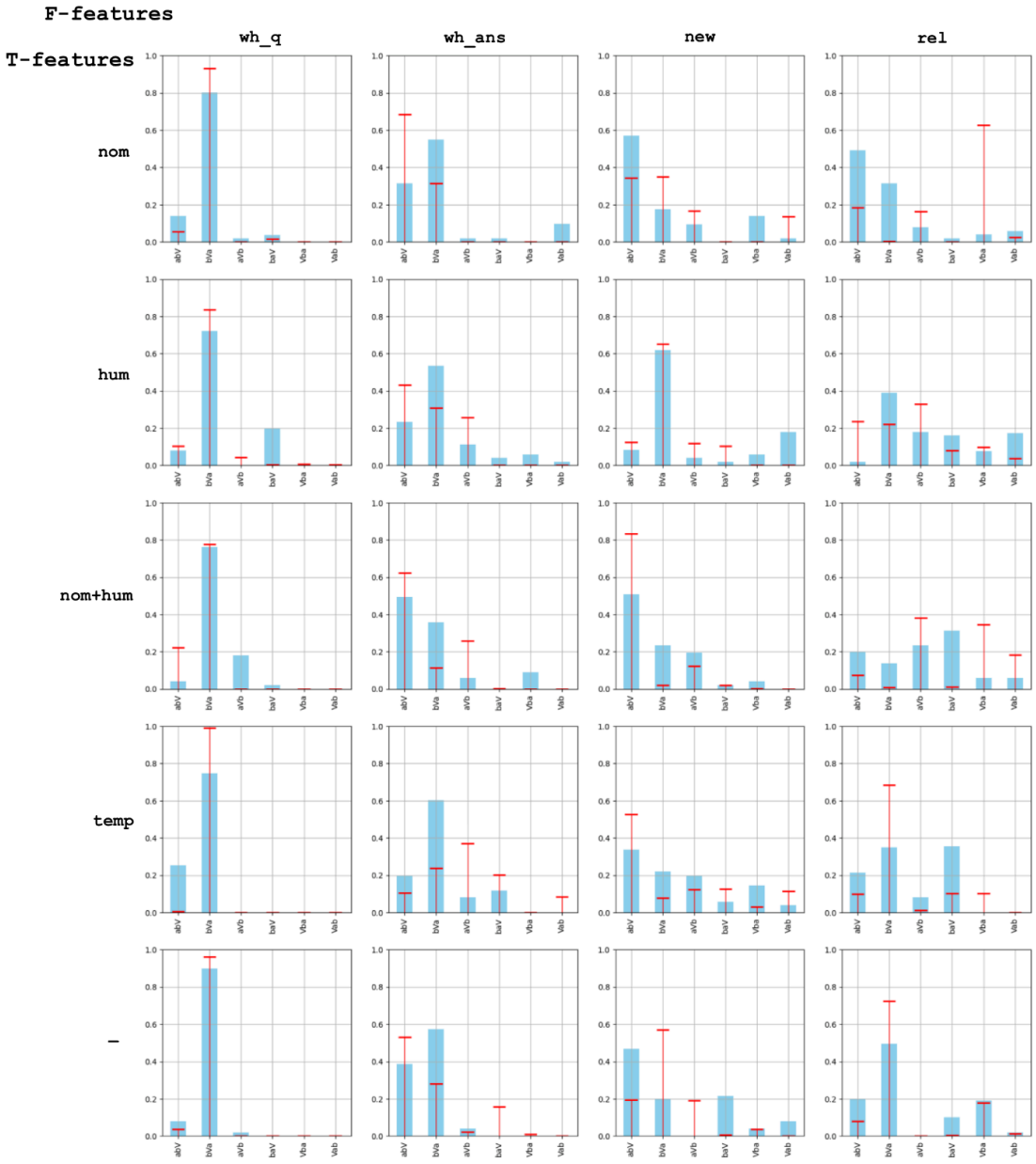

Figure F: normalized model probability assignments (red) mapped over survey relative ratings (blue)

4.2 KL divergence

To quantify this divergence between human and model preferences, I calculated the Kullback–Leibler (KL) divergence between the two distributions across all 20 scenarios. The overall KL-divergence was 0.7702, indicating a moderate distributional mismatch between the language model and human responses. However, this aggregate figure masks important differences across feature types, noted above. The following table shows the significant variation in KL-divergence based on column.

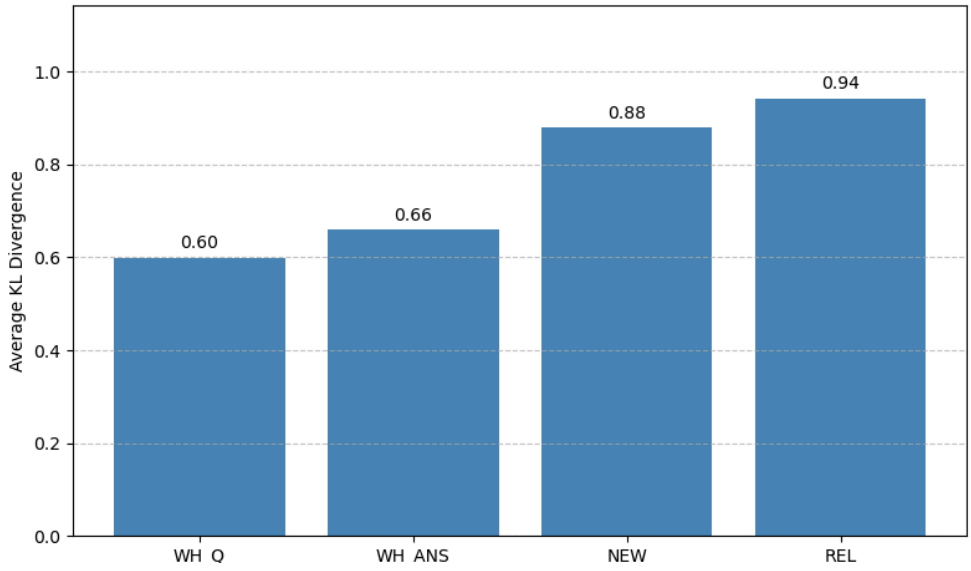

Figure G: variation in KL-divergence scores based on F-features

These values reinforce the observation that the model is more successful at approximating human judgments in contexts involving Wh-elements, but shows greater divergence in conditions involving relative clauses and informational novelty. This trend mirrors the findings of §3; the scenarios that elicited less consistent preferences from human participants are the same ones where the language model results diverge.

§5: Conclusion and Further Directions

5.1 Summary of results

Overall, the human survey results confirm that speakers of Hungarian do exhibit preferences in mapping discourse roles to syntactic positions. There appears to be a meaningful distinction between Focus-marking, where alignment between pragmatic function and syntactic position appears robust, and Topicalization, where there is less consistent concordance between function and position. However, within the scope of this idealized experiment, the difference between the two movement patterns was not found to be statistically significant. The language model output aligns partially with these human preferences, however with variation based on scenario type, with greater divergence in the contexts that also produced greater variability among humans. These results suggest that the discourse-configured orderings that are both well-seen in empirical data and well-modeled in language models (wh- question and answer pairs) are more obligatory, or perhaps categorically different from other cases of discourse configurationality, in which movement of particular roles into particular syntactic positions is not as obligatory and not consistently weighted higher by language model statistical calculations.

5.2 Assumptions and limitations

This study has aimed to explicitly compare human survey data on Hungarian word order preferences with probability assignments produced by a large language model. In doing so, it rests on the assumption—following Lau et al. (2016)—that linguistic knowledge is probabilistic in nature, and that human judgments of acceptability reflect probabilistic expectations over sentence forms. Crucially, Lau et al. distinguish between acceptability and likelihood of occurrence, pointing out that probabilistic models conflate surface-level frequency with

grammatical felicity unless properly constrained. However, the present study mitigates such confounds by holding constant sentence length, lexical content, and morphological marking across alternative word orders. This design choice isolates variation attributable to word order alone, justifying a direct comparison of both human and model responses to syntactic and pragmatic cues.

But this facet of the design also introduces a key limitation of the 100 constructed scenarios: all followed a uniform structure containing exactly three constituents. While this design allowed for controlled manipulation of discourse features and ensured consistency across conditions, it may have constrained the natural use of Focus and Topicalization. In real-world Hungarian discourse, these strategies may serve to highlight or disambiguate key information within more complex, multi-constituent clauses, by distinguishing salient elements from the rest of the postverbal material through movement. By limiting sentence length and syntactic complexity, the survey may have suppressed stronger preferences for Focus and Topic marking that would emerge more clearly in longer utterances with more constituents.

5.3 Suggestions for Improvement and Further Directions

The partial alignment observed between human results and language model results suggests that with targeted fine-tuning, the SambaLingo-Hungarian-Chat model may be capable of more accurately internalizing discourse-structural regularities. Fine-tuning might involve using explicit annotations of discourse features as training data. In particular, this would mean curating datasets that not only label syntactic categories but also explicitly specify information-structural roles like Topic and Focus for each constituent, allowing the model to learn mappings between discourse function and word order.

Another area for future improvement would be the incorporation of contrastive Focus, which was not explored in the current set of stimuli. Contrastive Focus structures, as in the following sentence, highlight a constituent by presenting it in explicit opposition to a salient alternative:

(8) *Nem <u>az apá-d-at</u> hív-tam meg, hanem az <u>anyá-d-at</u>.*
Not the father-POSS.2SG-ACC call-PST.1SG COMPL but.rather the mother-POSS.2SG-ACC
'It wasn't your <u>father</u> I invited, but your <u>mother</u>.'

Contrastive focus has been suggested to involve a different kind of prosodic emphasis and possibly make use of a distinct syntactic position (see É. Kiss 1998). Including such constructions in future surveys would shed further light on trends in word order preferences in Hungarian. A similar experiment could be designed by constructing minimal pairs of discourse scenarios that either license neutral Focus or require contrastive Focus, and surveying speakers' word order preferences across these pairs. Collecting acceptability judgments or forced-choice preferences across these two types would allow for testing whether models not only recognize Focus per se but also distinguish between neutral and contrastive Focus structurally.

In addition to expanding experimental stimuli, more robust findings could also be obtained through corpus analysis. A corpus study could examine Hungarian speech and written texts for naturally occurring instances of Focus and Topic marking, rather than a designed set of simplified scenarios. Such a corpus-driven approach would also allow for testing whether the patterns observed in short, constructed sentences generalize to longer and more syntactically complex utterances, potentially strengthening the empirical foundation for evaluating language model performance.

5.4 Concluding Notes

This research invites reflection on what kinds of errors matter when evaluating language models. While some deviations from human judgments may be judged irrelevant for language model evaluations, those involving violations of strong syntactic-pragmatic constraints result in utterances that are perceived as awkward or even ungrammatical. In asserting the importance of evaluating language models' ability to emulate human judgments about subtle syntactic shifts, this project follows the line of thought of Warstadt & Bowman (2022), who argue that evaluations of LLMs must move beyond coarse acceptability measures to capture subtle gradations of syntactic and pragmatic appropriateness. Work by Mueller and Linzen (2023) further supports the value of evaluating language model competence on subtle syntactic phenomena; their work demonstrates that even large models often fail to generalize over constrained syntactic environments like island structures—offering a cautionary note against equating high scores on simpler proficiency tests with deep grammatical competence.

More attention should also be paid to evaluating and improving language models in underrepresented languages. Most benchmark datasets and fine-tuning efforts have focused overwhelmingly on English and other "World Languages," leading to gaps in the models' ability to capture typologically diverse linguistic phenomena (Joshi et al., 2020). Prioritizing languages like Hungarian allows researchers to test models against complex syntactic-pragmatic mappings that are less common in English but prevalent cross-linguistically. Extending this type of analysis to other discourse-configurational languages such as Turkish, Korean, or Basque would further test the cross-linguistic generalizability of both human patterns and language model behavior, and contribute to a more comprehensive understanding of the linguistic capacities and limitations of current models.

Lastly, the methodological framework developed in this thesis contributes to ongoing conversations about empirical LLM evaluation. By triangulating theoretical prediction, experimental data, and model outputs, this work proposes a diagnostic paradigm that leverages the specificity of linguistic theory to test the limits of computational models. By systematically constructing Hungarian discourse scenarios and comparing native speaker judgments to language model outputs, the project provided both empirical insights on theoretical hypotheses about Hungarian word order and a novel framework for evaluating model sensitivity to discourse-structural regularities. In this way, the approach taken here parallels that of Zhu et al. (2025), who propose anaphora accessibility as a diagnostic for discourse-level understanding and highlight the divergence between human and model behavior when structural abstraction is required. Hungarian discourse configurationality, long studied as a window into the syntax-pragmatics interface and a means of testing assumptions of the generative tradition, also may through further scholarship prove valuable as a diagnostic tool for probing model alignment with human discourse knowledge. In this way, the project not only advances our understanding of a particular linguistic phenomenon but also serves as a proof of concept, highlighting the sorts of research methodologies that computational linguists might take on in the search for more precise language model diagnostics.

Works Cited

Beaver, D. I., Roberts, C., Simons, M., & Tonhauser, J. (2017). Questions Under Discussion: Where Information Structure Meets Projective Content. *Annual Review of Linguistics*, *3*(1), 265–284.

É. Kiss, Katalin. (1995). 'Introduction.' In K. É. Kiss, ed., *Discourse Configurational Languages*, Oxford Studies in Comparative Syntax. Oxford University Press.

———————. (2002). *The Syntax of Hungarian.* Cambridge University Press.

Grice, H. Paul, 1975: "Logic and Conversation." In Davidson, Donald, *The Logic of Grammar*, pp. 64-75: Dickenson Pub. Co.

Horvath, Julia. (1981). *Aspects of Hungarian Syntax and the Theory of Grammar* [Dissertation]. University of California, Los Angeles.

———————. (1986). *FOCUS in the Theory of Grammar and the Syntax of Hungarian*. Foris.

———————. (1995). 'Structural Focus, Structural Case, and the Notion of Feature-Assignment.' In K. É. Kiss, ed., *Discourse Configurational Languages*, Oxford Studies in Comparative Syntax. Oxford University Press.

———————. (2005). 'Is "Focus movement" driven by Stress?' In *Approaches to Hungarian 9: Papers from the Düsseldorf Conference*, Budapest, 131-58.

Jackendoff, Ray. (1972). 'Chapter 6: Focus and Presupposition,' in *Semantic Interpretation in Generative Grammar*. MIT Press.

Lambrecht, K. (1994). *Information structure and sentence form: Topics, focus, and the mental representations of discourse referents.* Cambridge University Press.

Joshi, P., Santy, S., Budhiraja, A., Bali, K., & Choudhury, M. (2021). *The State and Fate of Linguistic Diversity and Inclusion in the NLP World*.

Lau, J. H., Clark, A., & Lappin, S. (2017). Grammaticality, Acceptability, and Probability: A Probabilistic View of Linguistic Knowledge. *Cognitive Science*, *41*(5), 1202–1241.

Mueller, A., & Linzen, T. (2023). *How to Plant Trees in Language Models: Data and Architectural Effects on the Emergence of Syntactic Inductive Biases*.

Nguyen, T., Nguyen, C. V., Lai, V. D., Man, H., Ngo, N. T., Dernoncourt, F., Rossi, R. A., & Nguyen, T. H. (2023). *CulturaX: A Cleaned, Enormous, and Multilingual Dataset for Large Language Models in 167 Languages*.

Roberts, Craige (1998). 'Information structure in discourse: Towards an integrated formal theory of pragmatics.' *Semantics & Pragmatics* Volume 5, Article 6: 1–69.

Rounds, Carol (2009). *Hungarian: An Essential Grammar (2nd ed)*. Routledge Essential Grammars.

Szendrői, Kriszta. (2017). 'Focus Movement.' In *The Wiley Blackwell Companion to Syntax (2nd ed)*. John Wiley & Sons, Inc.

Warstadt, A., & Bowman, S. R. (2024). *What Artificial Neural Networks Can Tell Us About Human Language Acquisition*.

Zhu, X., & Frank, R. (2024). LIEDER: Linguistically-Informed Evaluation for Discourse Entity Recognition. In L.-W. Ku, A. Martins, & V. Srikumar (Eds.), *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 13835–13850).

Zhu, X., Zhou, Z., Charlow, S., & Frank, R. (2025). *Meaning Beyond Truth Conditions: Evaluating Discourse Level Understanding via Anaphora Accessibility*.

Zubizarreta, Maria Luisa. (1998). *Prosody, Focus, and Word Order*. MIT Press.